TECHNICAL REPORT TR 78-6-72

ON THE CREDIBILITY OF ESTIMATES: ITS EVALUATION AND IMPROVEMENT

DECISIONS AND DESIGNS INCORPORATED

Rex V. Brown

June 1978

ADVANCED DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making. These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics. The program is sponsored by the Cybernetics Technology Office of the Defense Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research – Engineering Psychology Programs. Participants in the program are:

Decisions and Designs, Incorporated
Harvard University
Perceptronics, Incorporated
Stanford Research Institute
Stanford University
The University of Southern California

Inquiries and comments with regard to this report should be addressed to:

Dr. Martin A. Tolcott
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

or

Dr. Stephen J. Andriole
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

9 14

TECHNICAL REPORT TR-78-6-72

6

ON THE CREDIBILITY OF ESTIMATES: ITS EVALUATION AND IMPROVEMENT.

by

10 Rex V. Brown

Sponsored by

Office of Naval Research
Engineering Psychology Programs
15 ~~Contract Number~~ N00014-75-C-0426

and

Defense Advanced Research Projects Agency
Contract Number N00014-78-C-0100

12 36p.

11 June 1978

DECISIONS AND DESIGNS, INC.
Suite 600, 8400 Westpark Drive
P.O.Box 907
McLean, Virginia 22101
(703) 821-2828

SUMMARY

The practical problem of appraising the accuracy of estimates--before or after they have been obtained--is analysed. A procedure called decomposed error analysis is proposed, which takes quantified assessments of different kinds of error, such as random sampling fluctuations and mismeasurement, and synthesizes them into a global assessment of error. It replaces and enlarges classical statistical inference approaches in a personalist format which does not depend on Bayesian updating. Applications from the private and public sector are presented.

CONTENTS

FIGURES

ON THE CREDIBILITY OF ESTIMATES: ITS EVALUATION AND IMPROVEMENT

1.0 INTRODUCTION

1.1 The Problem

Every decision maker, whether in business, government, or some other type of organization, relies on estimates of various kinds as a basis for resolving practical problems. Public policy and opinion are commonly based on mysteriously precise estimates of quantities whose magnitude cannot conceivably be known with any but the vaguest precision. As the Time Essay of August 2, 1971, "Of Imaginary Numbers," comments:

> From solemn public officials and eager corporations, from newspapers, television (and even, some dare say, from newsmagazines) comes a googol of seemingly definitive and unarguable statistics. They tell us, with an exactitude that appears magical, the number of heroin addicts in New York and the population of the world. By simulating reality, they assure us that facts are facts, and that life can be understood, put in order, perhaps even mastered.
>
> If this sounds fanciful, consider a few specimens from one issues of the New York Times last week:
>
> BANGKOK: In 1965, only 17% of the people in northeastern Thailand were within a day's journey of a main road. Today the figure is 87%.
>
> NEW YORK: The St. Patrick's Day parade cost the city \$85,599.61, where Puerto Rico Day cost only \$74,169.44.
>
> ATLANTA: There are 1.4 million illiterates in the U.S.
>
> KABUL: Caravans traveling between Afghanistan and Pakistan "commonly carry up to 1,200 pounds of opium at a time."

> If every statistic were regarded with...skepticism, it might well be found that many of our most widely accepted figures are..., at least in part, imaginary numbers. The national rate of unemployment, for example, is now stated to be 5.6%, but that figure is based entirely on people who officially reported themselves out of work. Idle students, housewives who cannot find outside jobs, unsuccessful artisans--such people are not counted. Statistics on crime are equally uncertain, since they mainly reflect police diligence in rounding up minor offenders and reporting all arrests.

Such estimates as noted above may be derived from formal research, notably by sampling or counting, from direct observation, or from hunch or "feel." Most commonly they involve a mixture of these sources of fact and opinion. However they may be derived, all estimates are subject to varying degrees of error. Thus, the decision maker--and the staff specialists who assist him--must somehow take account of the nature and extent of the errors associated with any estimate.

As far as we are aware, no serious, or at least widespread, effort has been made by presumably responsible purveyors of public or private estimates to so much as indicate "credible limits" on their estimates, let alone to seek reasonable grounds for such limits.

The reader may be quick to point out that ever since sample surveys came into widespread use, beginning in the 1920's, formal methods, notably concerned with confidence intervals, have been employed to appraise errors. However, they have almost invariably addressed only one class of errors, those arising from sampling fluctuations. Because sampling errors can be analyzed readily and explicitly, it sometimes appears that researchers treat sampling as the only source of errors in estimates. Experienced researchers and users of research results alike know that in most cases, "non-sampling" errors are much larger than pure sampling fluctuations. The unwary may be led to believe that estimates are far more precise than they actually are.

What has been lacking is a systematic method for analyzing total error in estimates, including errors arising from measurement and other sources as well as sampling error. This paper describes and explains one approach to the problem of evaluating total error. The result is not a complete, tested set of procedures; but it may be a useful step toward a very important goal. The results may interest managers, analysts, and research specialists in a variety of fields.

The orientation of this report is strictly practical, in the sense that the ultimate beneficiary is intended to be the man of affairs, a decision maker who may use the fruits of the statistician's labors rather than the statistician himself (i.e. the orientation is to contribute to the technology of administration and other applied arts).

The problem of analyzing total error is familiar to anyone who has to make decisions in the face of uncertainty, and consists of two parts:

1. how to assess uncertainty about relevant target variables (such as a market share), which we will call the problem of target assessment;

2. how to evaluate ways of reducing this uncertainty, which we will call the problem of research design.

1.2 Illustrations of Target Assessment and Research Design Problems

When a policy maker or executive looks at a completed piece of research relating to some target variable, such as the military strength of a potential adversary for a defense official, or the demand impact of an advertising campaign for a businessman, he will normally have two questions in mind: What should his own "best" estimate be? How much faith should he have in this estimate?

If the executive is not interested implicitly in questions along these lines, then it is not at all clear how the research can have a bearing on his decision making, or why the research was undertaken in the first place. (Organizational prestige, the relief of personal anxiety, or the desire to sell a decision already made, are not unknown motivations for research, of course!) How he does or should resolve such questions is open to question (but not arbitrary choice). He may adopt as his own best estimate whatever raw number (estimate) is thrown up by the research (in a business setting, if 5% of widget users surveyed claim to use brand X, 5% would be his estimate of the national brand share). Alternatively, he may want to adjust that raw estimate in the light of any prior views he may have of the research technique used or the target variable itself.

As far as the executive's faith in his best estimate is concerned, he may treat the estimate as a certainty in his subsequent thinking; or, he may use some "objective" statistical procedure to set a "confidence interval"; or, he may somehow take account of his personal judgment in assigning a margin of error or, as it is technically called, a credible interval.

Intuitively, hard-headed administrators make their own assessments and adjustments all the time, without recourse to a theoretician. A business executive, for example, would be quite likely to make an informal research appraisal of the following kind: "This report says our company has 5% of the widget market. Ridiculous! We are selling 5,000 a week and the total market cannot be more than 50,000. Probably some of our customers in the survey said they bought the competitor's brand because he advertises more. I would up that estimate to 10% give or take a few percent."

Defense officials evaluating intelligence reports will often, and with good reason, make similar responses. Estimates of interest might include the throw weight of a Soviet missile, the number of Soviet troops stationed in Poland, the proportion of Soviet aircraft equipped with certain advanced fire control systems, or the number of new Soviet tanks in East Germany.

1.3 The Need for Formal Aids

Now it is quite possible for an executive to do a perfectly good job of combining survey evidence with his experience in making such an appraisal by using no more than his informal common sense. On the other hand, he may welcome some formal assistance in weighing the evidence.

A realistic appraisal of the accuracy of an estimate will clearly help a decision maker to use that estimate effectively. It may also provide a useful stimulus to improving the estimating process itself. It is a familiar phenomenon that appropriate measures of effectiveness for any task (like estimation) tend to improve the performance of that task. Who can doubt that Neilsen ratings have had the effect (however deplorable) of moving TV programs in a direction which maximizes the number of sets turned on (which, of course, is what Neilsen measures)? The fact that the accuracy of election polls can be checked quickly and surely no doubt accounts, in large measure, for the high degree of accuracy of such polls. If we can measure, however tentatively, the accuracy of estimates used in public or private sectors, perhaps this will put effective pressure on the researcher (estimator) to make his estimates more accurate.

In defense and related areas, as in business, quantitative research is almost invariably action oriented, and the case for a meaningful and comprehensive evaluation tool addressing user interest becomes irresistible. The case can perhaps be made (suspect, in my opinion) that scientific research should only be reported in classical terms, i.e. restricting attention to objectively measurable sources of error, like random sampling. Surely no case can be made for so restricting the appraisal of estimates on which national policy may be based.

In current practice in military intelligence, classical validation tests are in fact used in only a small proportion of cases involving quantitative estimates. Such tests are limited by the number of people qualified to apply them and are normally performed only by scientific specialists. Intelligence analysts typically are not trained in these methods. They may make statements of the form "such and such Warsaw Pact Division has 8,500 men in it plus or minus 10%." Such an assessment would take into account all the considerations the analyst thought relevant, but it would be presented without formal validation for the latter interval or an indication of how probable it is that the true number lies within that range.

Because of the requirement (actual or perceived) that quantitative estimates be publicly documented, it is not uncommon for private estimates made by analysts to differ from those made public. The latter may only consist of elements that can be firmly defended, and the former may include richer but more diffuse and less readily verified and validated data in which they nonetheless have more confidence. It is not apparent whether any consistent bias exists between the two. However, the private, more realistic estimate will typically be hedged by a larger margin of uncertainty than the public estimate.

Outside of the military, other government agencies also engage in making estimates and designing surveys. For example, the Federal Energy Administration is currently concerned with how to specify data-gathering projects in order to produce the most credible estimates, allowing for biases and other sources of error in estimates received. Such surveys will seek estimates of:

- financial and other operations of oil companies;
- the availability of natural gas supplies at peak demand times; and
- demand patterns of motorists and other energy consumers.

Finally, the need for user-oriented appraisal of estimating strategies and estimates is nowhere more evident than in the social and natural sciences, whose empirical core is based on experiments and other sample inquiries. The conventional scientific validation procedures of classical statistics, such as specification of confidence intervals and tests of significance, are partial and typically confusing measures from the user's point of view, useful as they may be for standarized documentation of reported experiments (see Section 2.1 below).

It would be useful, therefore, for government agencies which carry on a significant amount of social science research to have a more complete and less confusing method of appraising estimating strategies and estimates.

The techniques discussed in this paper have been used in particular by the Federal Energy Administration to estimate conservation behavior of households and the market for solar heating devices in the home as presented in Brown et al. (1977) and Campbell et al. (1977). A fuller development appears in Brown (1969).

2.0 CURRENT STATE OF THE ART

2.1 Classical Inference Techniques

Of course, the technical literature abounds with procedures that appear to address problems such as those mentioned above. When sample findings are available, for example, common devices such as confidence intervals, maximum likelihood estimates, and tests of significance certainly seem to be saying something about what we call target assessment. However, the trouble with classical inference tools such as these is that their output is not in a form that is of direct interest to a decision maker. He wants to answer the very personal question, "Where does my target variable probably lie?", whereas a confidence interval, for example, is telling him how surprising the observed sample would be if some variable (not necessarily his target variable) had some hypothetical values. This simply is not answering any question a typical executive would want to ask.

For example, the confidence interval says something very difficult for the layman to interpret (or use, if he can interpret it), as follows. "Intervals calculated as this one was from repeated samples will include the true value 95% of the time." A special, but common, case of a 95% confidence interval is computed as follows: The lower limit is selected such that, if it were the true value of the target variable, repeated sampling would produce a research statistic larger than that actually obtained 2-1/2% of the time--and conversely for the upper limit. Figure 2-1 gives a graphic illustration, based on a simple random sample of nine hundred, of which ninety showed the property in question. (The population from which it is drawn is effectively infinite.)

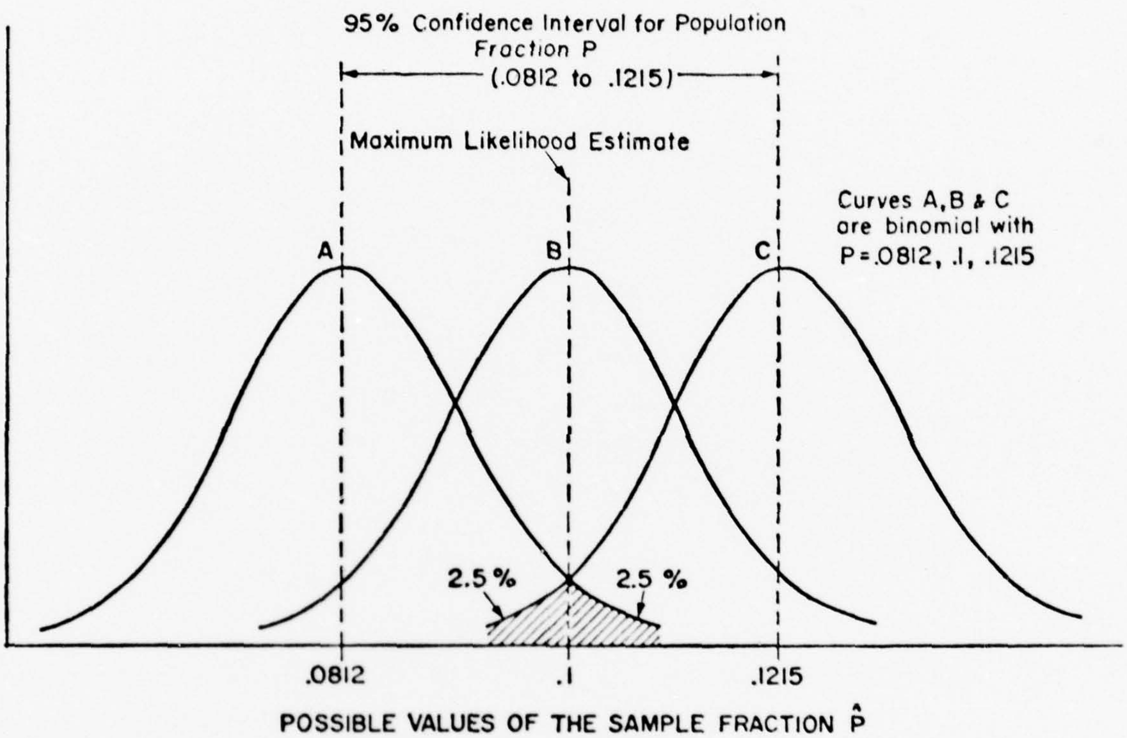


Figure 2-1
CLASSICAL APPRAISAL OF SURVEY FINDINGS

Where the target variable is a fraction p, for example, and a sample of size n produces a fraction $\hat{p}$, the 95% confidence limits are approximated by

$$\hat{p} \pm 2(\hat{p}(1 - \hat{p})/n)^{1/2}$$

and this formula is in very common use.

Substantial literature has been developed on the considerable variety of confidence interval techniques in use. However, all of them partake of the same general character already discussed and which will be developed next. The differences are not critical to this inquiry.

2.2 Classical Inference Applied to Camford Case

In the real study on which the Camford example is based, classical inference was attempted in a way which is very typical in survey estimate appraisal. In the sample of nine hundred locally registered car owners, it will be recalled that ninety, or 10% of the sample, reported that they would park at peak hours on the given days, if meters were introduced. Approximate 95% confidence limits computed according to the formula just presented are

$$.1 \pm 2(.1 \times .9/900)^{1/2} \text{ or } 8\% \text{ to } 12\%$$

which is what appeared in an Appendix to the original Camford report. The exact limits, computed by a computer program are 8.12% to 12.15%. Figure 2-1 shows how these inferences are built up: 10% is a "maximum likelihood" estimate, in the sense that 10% is more likely to be the sample value, if 10% were the true fraction in the sampled population, than if the population had any other fraction.

The three curves are sampling distributions showing the probability of obtaining any particular sample fraction given the true fraction of population being sampled (not necessarily the population of interest). If the lower limit A were the true fraction, repeated samplings would produce a fraction larger than that actually obtained 2-1/2% of the time. The reverse is true for C, the upper limit.

Now at first sight, it might appear that the target assessment questions posed earlier have been answered. Indeed, a large fraction of the countless users of confidence limits would have the impression that:

(1) 10% is the "best" single estimate for the true proportion of "metered parkers" in the frame of local motorists sampled;

(2) it is reasonable to assign about 95% probability to the true proportion lying between 8% and 12%.

In general, neither interpretation can even approximately be supported (see Brown 1969, pages 73-82).

When an executive considers the research design (as opposed to target assessment) problem, he wants to answer a question like "What research can I do which will make me least uncertain?" It would never occur to him, and rightly so, to ask, "What research will produce the smallest sampling variance from among those research designs for which a sampling variance can be objectively calculated?" The latter is the kind of information he might extract from the currently dominant tools of classical inference.[1]

---

1 A more general and technical discussion of the weakness of classical inference for decision-making purposes appears in Pratt _et al_. (1965), Chapter 20.

Instead, when assessing his target variable, the decision maker surely wants to come up with a personal probability assessment, possibly in the detailed form shown in Figure 2-2. In most cases, he will be satisfied with a simple summary of the distribution, say, as an interval within which he is 95% sure the target variable really lies--in this case 300 to 2200--or possibly just his expectation--in this case, 1100.

Similarly, when choosing among research designs, he will want to look ahead to the kind of probabilistic assessment he can expect to make after the research. Presumably, he will opt for the design which, in some sense, promises to produce a personal probability distribution with as little "spread" as possible.

2.3 Personalist Decision Analysis

A new branch of statistics known as <u>personalist decision analysis</u> (PDA) is available to handle personal decision and inference problems of this kind.[2] Specific variants known as Bayesian probability updating and preposterior analysis have been substantially developed to address exactly these problems.[3] However, even though military analysts and other staff people have been exploring the applications of these specific tools,[4] they have been slow to take hold among real-world decision makers. A survey of business applications of PDA[5] found very few instances where executives <u>acted</u> on the implications of such analyses (though plentiful use of other variants of PDA, notably decision trees, was reported).

---

[2] See Savage (1972), Raiffa & Schlaifer (1961).

[3] See Brown <u>et al</u>. (1974), Chapters 35, 26.

[4] See Barclay <u>et al</u>. (1977), Chapters 4 and 5.

[5] See Brown <u>et al</u>. (1974), Chapter 7.

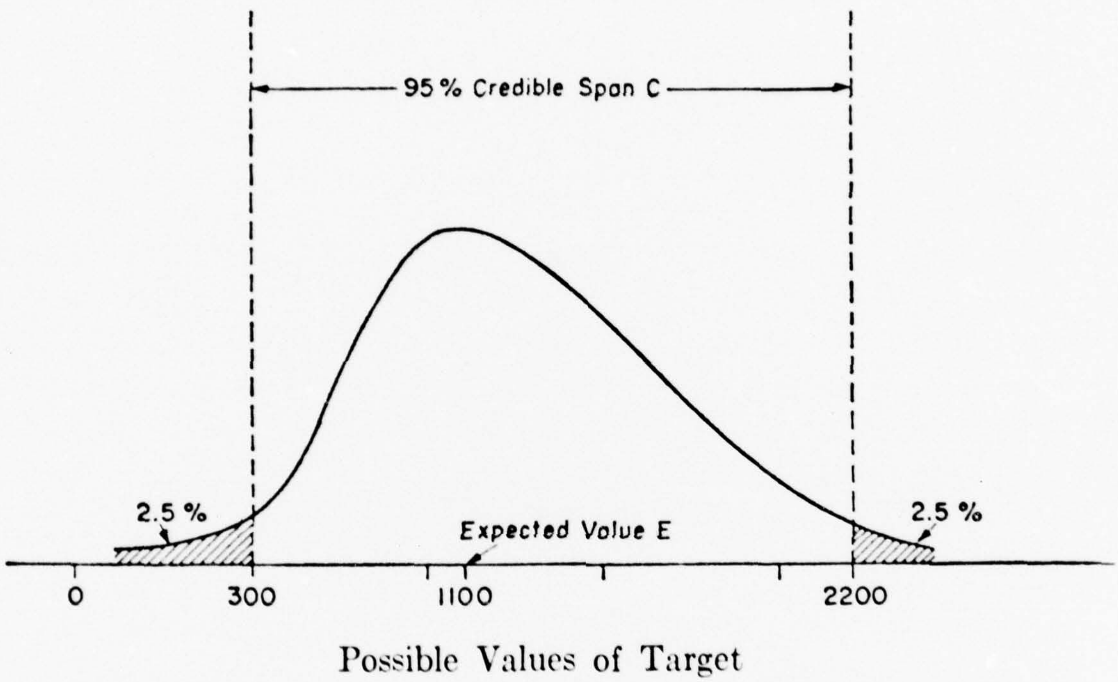


Figure 2-2
PERSONAL PROBABILITY FOR TARGET VARIABLE

No doubt, part of this lack of implementation is due to the quite natural lag between a new technology's development at a theoretical level and its becoming operational. But, part may be because the technique itself, as currently developed, is not always appropriate for use by nontechnical decision makers or executives.

In particular, the technique involves Bayes' Theorem and requires the executive to participate in the assessment of esoteric inputs (e.g. "prior distributions" and "likelihood functions") which his training typically does not equip him to supply or even to understand.

Moreover, very few executives feel that they understand even the general purpose of these devices. For this reason, they are understandably hesitant to trust decisions that may involve millions of dollars of private and public resources to an analysis based on an arcane logic.

Is there any way of avoiding these drawbacks? We feel there is and propose an alternative which, while it is personalistic in the sense that it accepts personal inputs and its output is interpreted personally (like the tools just mentioned,) it does not depend on Bayes' Theorem (which they typically do) and hopefully avoids some of its drawbacks.

3.0 A SUGGESTED APPROACH

3.1 Decomposed Variable Analysis (DVA)

The decomposed variable analysis[6] technique depends not on Bayes' Theorem, but on the equally well-known logic of the distribution of functions of random variables. In the special version to be presented, it can be used both for problems of target assessment and of research design.

The essential steps are very simple and are as follows:

1. The target variable is decomposed, in the sense that it is expressed as a function of two or more components. A very simple example would be to express future demand for energy as energy per consumer times number of consumers. A slightly more elaborate decomposition (and decompositions can get very elaborate) would be to express energy as the sum of multiplicative expressions of the above form for each of a number of use sectors, such as lighting, heating, transportation, etc.

2. Each component thus defined is assessed probabilistically (e.g. in the form of a personal probability distribution) on the basis of whatever evidence is available to the assessor. This evidence could include field work, judgment, or published statistics, and the supporting reasoning could be any combination of intuition and statistical theory (including possibly, Bayesian probability updating).

---

[6] Also referred to as "credence decomposition," e.g. in Brown (1969).

3. A personal probability distribution (e.g., in the form of Figure 2-2) is derived routinely by standard statistical procedure from the component distributions and the decomposition formula by which they are combined. Computer programs, mathematical formulas, and other supporting devices have been developed to make this processing as painless as possible.

For example, when a target variable, t, is decomposed into a product of components (for example, $t = x \cdot y \cdot z$), the required distribution for t can be approximated from assessed distributions for the components x, y, and z as follows.

Let the mean of t be represented as E(t), and let the 95% credible span of t be represented as C(t). (The assessor assigns a 95% probability that the target variable lies within the range specified as the 95% credible span.) Assess for each component the mean (E(x), and so on) and credible span (C(x), and so on). Then, if the components are judgmentally independent of one another, or nearly so, the follow approximations hold.[7]

$$E(t) = E(x) \times E(y) \times E(z)$$

$$C(t) = E(t) \times \sqrt{\frac{C(x)^2}{E(x)^2} + \frac{C(y)^2}{E(y)^2} + \frac{C(z)^2}{E(z)^2}}$$

For example, suppose the "min," mean, and "max" (note that "min" and "max" refer to the edges of a 95% credible

[7] As explained in Brown (1969), Chapter 9.

interval, not absolute limits) of the three components are assessed to be:

x: 5,10,15

y: 10,60,120

z: .1,.5,.8

Applying the above formulas will give:

$$E(t) = 10 \times 60 \times .5 = 300$$

$$C(t) = 300 \sqrt{\frac{10^2}{10^2} + \frac{110^2}{60^2} + \frac{.7^2}{.5^2}}$$

$$= 300 \sqrt{1 + 3.36 + 1.96}$$

$$= 300 \sqrt{5.32}$$

$$= 692$$

If the distribution of t were symmetrical, the "min," mean, and "max" would, of course, be given by 300 ± 346. We can usually get a better approximation to the edges of the credible interval by assuming "log-symmetry," which implies that the "upper edge" divided by the mean equals the mean divided by the "lower edge." This is still only an approximation, however, and is arithmetically bothersome. (The exact determination would require more detailed input and statistical theory.) One can do just about as well by locating the credible interval by eye, viz.:

t: 120, 300, 812

(Note that 812 - 120 = 692.)

The above routines are approximations, but adequate as a first pass for many real problems. (A method for obtaining greater precision and generality, for example by using simulation, is discussed in Section 9.2 of Brown (1969). Specific computer programs have been developed for this purpose at Decisions and Designs, Incorporated.)

3.2 Decomposed Error Analysis

At this level of generality, decomposed variable assessment is a rather trivial (if grossly under-exploited!) tool. However, there is a variant of DVA, decomposed error analysis (DEA), which is less obvious and which seems to lend itself rather conveniently to problems of target assessment and research design.

In DEA, what is decomposed is not the target variable of ultimate interest to the executive but rather the estimating error resulting from a specific piece of quantitative research, such as a sample survey.

There are at least two ways of formally defining estimating error. It can be defined as the difference between the target variable and some more or less arbitrary estimate calculated from the research findings. Alternatively, it can be the ratio of the target variable to such an estimate. Either formulation has advantages in particular circumstances, though for illustrative purposes only the error ratio will be discussed.

3.2.1 An urban planning example - The British town of Camford had a mail survey done to assess the probable demand for parking space if meters were introduced. A list of ten thousand locally registered motorists was obtained, and one thousand were randomly selected and sent questionnaires. Nine hundred returned the questionnaires and, of these, 10% or ninety, indicated that if meters were introduced, they would be parked in the downtown area at a given peak hour.

The city engineer's target assessment problem is what to conclude about the actual demand for parking space if meters were introduced, expressed in a form like Figure 2-2. He also has a research design problem. If he conducts a new survey in another town, should he use the same budget on another mail survey or rely on a smaller personal survey?

3.2.2 DEA for target assessment - On the target assessment problem, the first thing the city engineer might do using decomposed variable analysis would be to decompose the total spaces needed (if meters were introduced) into the product of:

1. the fraction of local motorists needing space (t);

2. the number of local motorists (n); and

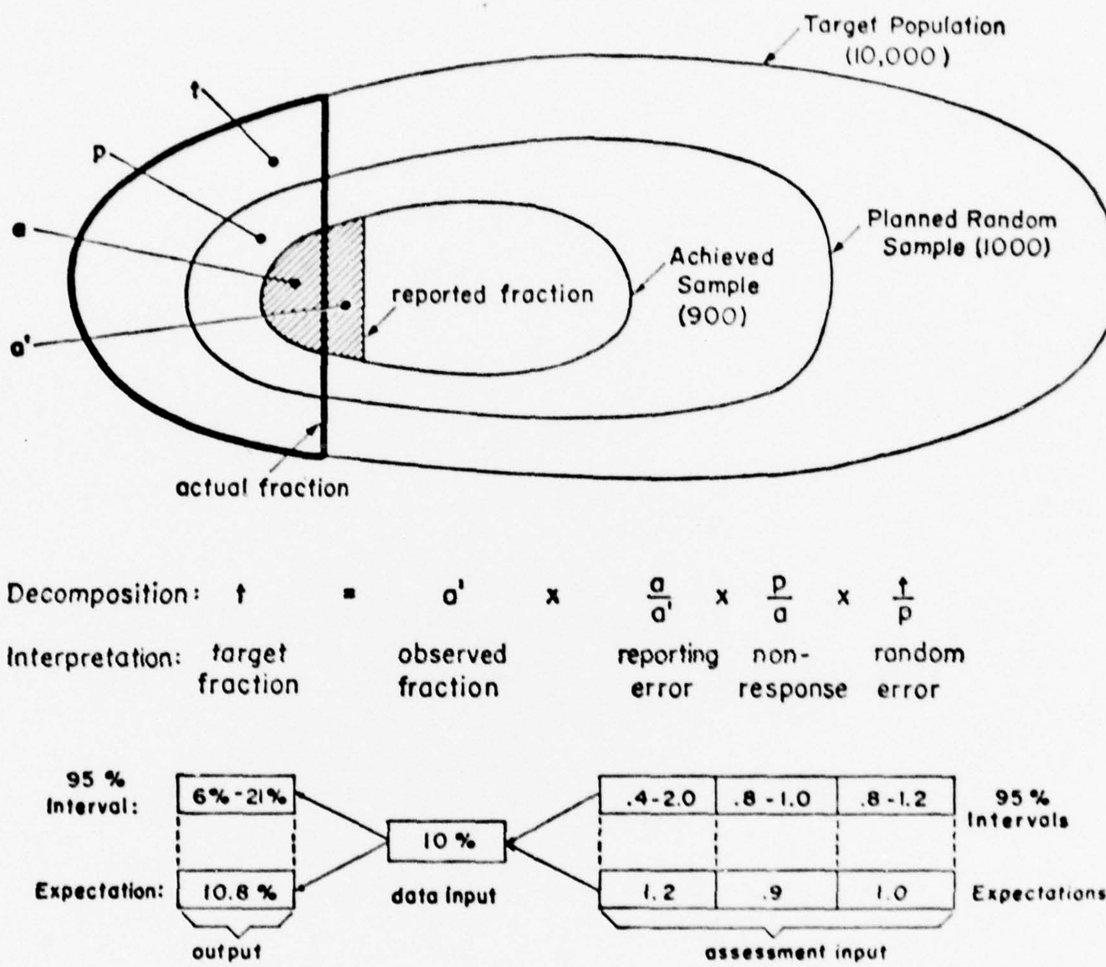
3. some adjustment factor intended to allow for spaces needed by out-of-town parkers (f).

If probabilistic assessments can be made for each of these, a probability distribution on the target variable can be derived routinely. The number of local motorists (n) is known to be ten thousand, so no probabilistic assessment is needed of that component. The out-of-town adjustment component (f) can be assessed informally by direct intuition. This leaves the "local fraction" (t), the variable which the mail survey addresses. The city engineer may have more misgivings about informally assessing a probability distribution on this variable, so he might decide to perform an error assessment of (t).

Figure 3-1 shows the essential steps the assessor might go through in order to express total error ratio as a function of component ratios which reflect distinguishable (and assessable) sources of error. The nested rings at the top of the figure and the vertical lines indicate the various ways in which sources of error can enter between the true value of the target, t (the "local fraction"), and the estimate, a' (known to be 10%).

Thus, t/a' is the total error ratio, and the component error ratios are defined in the line in Figure 3-1 marked "Decomposition." It can be seen that three sources of error are distinguishable: random error, nonresponse error, and reporting error. It can easily be verified that each error ratio will equal one if there is no error of that type involved. The set of boxes on the right hand side of the bottom line of Figure 3-1 summarizes, in the form of 95% credible intervals, probabilistic assessments that were made for each of three component error ratios. The detail of these assessments and the logic behind them are described in Brown (1969), Chapter 8.

The expectation and credible interval for the total error ratio t/a' are calculated from the component error assessments via the product decomposition formula given in Section 3.1 above as 1.08 and .6 to 2.1 respectively. Multiplying t/a' by a' (=.1) yields an expectation for the target fraction of 10.8% and a 95% credible interval of 6% to 21%, as shown in the left hand boxes of Figure 3-1.



NOTE: The top diagram defines the quantities in the Error Ratio Decomposition below it. The numerical input and output appear in the lower boxes below the corresponding elements in the decomposition.

Figure 3-1
ASSESSING A POPULATION FRACTION FROM A SURVEY ESTIMATE USING DECOMPOSED ERROR ANALYSIS

The city engineer might thus conclude, if he accepts the input assessments, that he can be 95% certain that the local fraction t lies between 6% and 21%, with an expectation of 10.8%. Conjoined with the knowledge that there are ten thousand local motorists and an assessment of the "out-of-town adjustment" with a credible interval of 1 to 1.2, a distribution on the real target variable, total spaces needed, was derived and is displayed in Figure 2-2. Therefore, his target assessment is that, with 95% personal probability, between 300 and 2200 parking spaces will be needed if meters are introduced.

3.2.3 DEA for research design - For the research design problem, the city engineer would go through virtually the same procedure for each of the alternative research designs considered. If the cost is the same, he might reasonably choose whichever strategy leads to least uncertainty, as measured, say, by the span of the credible interval on the total error ratio. Alternative research design criteria can be selected, such as prior expectation of posterior variance, but they seem to produce almost identical rankings.

It is possible that the most important applications of DEA will be not in appraising research estimates after the fact, but rather in choosing research strategies from which estimates will emerge.

The following research design applications of DEA are examples drawn from the author's experience. Although the contexts are largely business oriented (other than the last), analogies with research design problems in other areas can readily be made.

1. To estimate brand sales, should a consumer panel be used to estimate sales per consumer, or a store audit performed to estimate sales per store? In either case, total error depends on errors in estimating the size of the population (consumers or stores) and in estimating sales per unit. Decomposition and four separate error assessments helped decide that, in a particular instance, a store audit promised the least serious combination of errors.

2. To estimate annual replacement demand for shock absorbers, how should the decomposition be formulated? Should it be vehicles in circulation times annual replacement rate (estimated from information from vehicle manufacturers and motorist interviews)? Or, should it be the product of the number of garages times the average replacement sales per garage (requiring a garage survey)? Or, should both approaches be used and, if so, in what proportion? (The latter strategy was selected with the main emphasis on the second approach.)

3. Which of several sampling lists should be used? (The less complete list may contain classifying information which permits a more efficient sample design, but the omissions may have important distinguishing features.)

4. Should random or quota sampling be used to estimate family savings patterns? (Random

sampling promises a more representative sample, but quota sampling perhaps promises more believable respondents.)

5. To estimate the number of welding sets in use, should a simple random sample of industrial companies be preferred over a judgmentally skewed sample which favors large companies?

6. To establish how many automotive parts were purchased by a vehicle fleet operator, would you ask him or sample his maintenance records? (This permits a trade-off between convenience and accuracy.)

7. If several different estimates of the same target are available based on different sources of information, how should the data be pooled?

8. What is the right economic balance of research resources between gathering data and analyzing it?

9. In estimating energy demand, should many converging approaches be used and the results pooled or should all available resources be devoted to a single estimating approach? DEA suggested the former.

4.0 CONCLUSIONS

4.1 How Can the Policy Maker Use DEA?

The patient reader who has borne with us thus far and has been persuaded to try decomposed variable analysis on research design or on target assessment problems may wonder how, precisely, to proceed.

It is unlikely that the typical executive needs to involve himself in much more detail than is covered in this report provided he understands very clearly the input (assessments) and the output (conclusions) of the analysis. He may, however, wish to delegate the detail and/or confer with someone experienced in using the technique. In our experience, the greatest dangers inherent in any type of formal approach to executive problems, including operations research and other approaches in common use, are that the problem solved is different from the problem the executive has and that assumptions underlying the analysis are unacceptable to the executive (although he may not be aware of the assumptions). These are particularly serious complaints against conventional uses of statistics for research appraisal.

This is not to say that decomposed variable analysis invariably needs the participation of a technical specialist. If the decomposition of the target variable goes no further than a few intervening variables without explicit error decomposition for any one of them, DVA can be quick and trouble-free even for the layman.

Suppose an executive requires a quick but reasonable probabilistic assessment of some quantity of interest, and his information and judgment are based on varied and diffuse sources. He could decompose this target into a few components on which his experience independently bears, making direct intuitive assessments of each and using a simple formula (or computer program) to process them.

More specifically, suppose the target variable is the demand for gasoline at $1 a gallon in two years time. It can be decomposed as the product of:

1. how many motorists there are now;

2. the rate of growth of the motorist population over the next year;

3. the individual average mileage of the motorists; and

4. the average consumption of gas per mile.

The executive has then only to think about these components in turn and judgmentally assign an expected value and 95% credible interval to each. By applying a simple arithmetical procedure,[8] a "best forecast" and a credible interval for the target are quickly obtained.

4.2 The Appraisal Tool Appraised

While the approach proposed here may not be the best that can be devised, it does appear to offer substantial advantages to the research user over any of the alternatives he currently has at his disposal. Any moderately good appraisal technique which takes account of all major sources of error and which gets used is an improvement by an order of magnitude over current practice.

---

[8] See Section 3.1 above.

As George Kennan has said in a slightly different context, "Tentative solutions to major problems are worth more than definitive solutions to trivial problems." A major object of this paper will have been achieved if research users are encouraged to press for at least tentative solutions to the major appraisal problems they face and to be a little more suspicious of definitive but trivial appraisals which they all too commonly receive.

While decision makers concerned with clarifying their own uncertainities will surely support any move in the direction of realistic target assessments (especially if quick and cheap), resistance to progress can be expected from two quarters: researchers whose work will come under more stringent scrutiny and research commissioners who have an interest in "proving" something to third parties (for example, that their magazine penetrates markets attractive to advertisers).

It is up to the ultimate research user--the executive--to make sure that realistic target assessments are made, whether this approach or some other is used. Even if the user does not make the assessment himself, he can at least bring pressure to bear on the researcher to produce his own assessment in a form which makes the underlying component assessments explicit and, therefore, subject to review (though it is clearly more satisfactory to use an appraiser who is not beholden to the research practitioner). One objective of this study has been to dispose of the claim, previously tenable, that no operational and logically sound way of appraising total error exists for target assessment and that, therefore, no attempt needs to be made.

Though it is perhaps too much to hope that the research practitioner will carry out realistic target assessment himself (at least for public consumption), he will surely be motivated to try realistic research design appraisal, particularly if he ultimately expects his research estimates to be subjected to appraisal (by whatever means) of their credibility. In this way, the existence of at least one systematic scheme for appraising the credibility of estimates could conceivably lead to dramatic improvements in the practice of quantitative research.

Although the general decomposed variable analysis technique and the decomposed error variant of it show good promise in marketing and survey research where they have most frequently been applied, even there they are at a primitive state of operational development. Other researchers, notably Professor Charles Mayer of York University, have been working on the critical problem of how to make reasonable, empirically based component assessments which are required by these techniques or others with the same objectives.

Needless to say, a great deal of additional work needs to be done generally in the area of the credibility of estimates, for example, in ironing out operational bugs in specific techniques and in building a solid empirical base upon which to make required assessments.

REFERENCES

1. Barclay, S.; Brown, R.V.; Kelly, C.W. III; Peterson, C.R.; Phillips, L.D.; Selvidge, J.E. Handbook for Decision Analysis. McLean, Virginia: Decisions and Designs, Inc., 1977.

2. Brown, R.V. Research and the Credibility of Estimates. Boston, Massachusetts: Harvard University, Graduate School of Business Administration, Division of Research, 1969.

3. Brown, R.V.; Campbell, V.N.; and Repici, D.J. Analysis of Residential Fuel Conservation Behavior. McLean, Virginia: Decisions and Designs, Inc., October 1977. (Sponsored jointly by the Office of Naval Research and the Federal Energy Administration.)

4. Brown, R.V.; Kahr, A.S.; and Peterson, C.R. Decision Analysis for the Manager. New York: Holt, Rinehard and Winston, 1974.

5. Campbell, V.N.; Brown, R.V.; Repici, D.J.; and Rhees, T.R. An Attitudinal Study of the Home Market for Solar Devices. McLean, Virginia: Decisions and Designs, Inc., 1977.

6. Schlaifer, R. Introduction to Statistical Decision Theory. New York: McGraw Hill Book Company, 1965.

7. Raiffa, H., and Schlaifer, R. Applied Statistical Decision Theory. Boston, Massachusetts: Harvard University, Graduate School of Business Administration, Division of Research, 1961.

8. Savage, L.J. The Foundations of Statistics. Dover Publications, Inc., 1972.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies
Office of Naval Research Attention: Code 102IP 800 North Quincy Street Arlington, Virginia 22217	6 copies

Revised March 1978

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense Research and Engineering (Test and Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Assistant Director (Environmental and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: COL Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
APO San Francisco 96610

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Attention: R. J. Roland, (Code 52R1)
C^3 Curriculum
Monterey, CA 93940

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 5403)
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Alan H. Curry
Operations and Management Science Division
U.S. Army Institute for Research in Management Information and Computer Science
730 Peachtree St., N.E. (Suite 900)
Atlanta, Georgia 30308

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems Research Laboratory
U.S. Army Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office (Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development and Acquisition Programs
Office of the Deputy Chief of Staff for Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, DC 20330

Director, Advanced Systems Division (AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norma, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR 78-6-72	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON THE CREDIBILITY OF ESTIMATES: ITS EVALUATION AND IMPROVEMENT		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Rex V. Brown		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0426 N00014-78-C-0100
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decisions and Designs, Incorporated Suite 600, 8400 Westpark Dr. P.O.Box 907 McLean, Virginia 22101		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209		12. REPORT DATE June 1978 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Target assessment
Classical inference
Personalist decision analysis
Decomposed variable analysis
Decomposed error analysis
Target variable
Credible interval
Confidence interval

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The practical problem of appraising the accuracy of estimates--before or after they have been obtained--is analysed. A procedure called decomposed error analysis is proposed, which takes quantified assessments of different kinds of error, such as random sampling fluctuations and mismeasurement, and synthesizes them into a global assessment of error. It replaces and enlarges classical statistical inference approaches in a personalist format which does not depend on Bayesian updating. Applications from the private and public sector are presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)